

Assignment VII : Clustering Methods

April 16, 2025

ID5002W: Industrial AI Laboratory

Total Marks: 80

Instructions

1. Assignment shall be submitted before the due date. Late submissions will not be entertained. If you cannot submit the assignment due to some reasons, please contact Dr. Tirthankar by email.
2. All the assignments must be the student's own work. The students are encouraged to discuss or consult friends or classmates. However, they have to submit their own work. Any malpractice will be reported to the authorities and action will be taken as per the IIT Madras rules.
3. If you find the solution in the book or article or on the website, please indicate the reference in the solution.
4. You are expected to submit “*.py” file instead of notebook.
5. Import the given data in “*.npz” format.
6. A short report in pdf format along with the code should be submitted containing results and analysis as asked in the assignment.
7. Please note:
 - Code should execute without any error.
 - The code should be clean with readable comments.
 - The output of the code should be in a relevant format so that it can be understood by an evaluator.
 - For figures, **do not** use `plt.show()`. Instead, use an appropriate command to save the figure, such as: `plt.savefig("roll_no_figure1.png")`. The evaluator should understand the figure content from the file name. *The entire string should be in lower case.*
8. Grading Policy:
 - 50% code correctness.
 - 10% code readability and comments.
 - 40% on report and analysis.

Problem 1

Given a dataset of points arranged in three distinct, non-overlapping concentric circles, implement a clustering strategy to identify and separate the inner and outer parts of each circle. You are provided with a 2D dataset. The data points visually form three separate groups, each consisting of an inner and an outer circular distribution.

- (a) Implement a two-stage clustering approach:
 - Stage 1: Group Identification: Apply a suitable clustering algorithm to identify the three distinct circular formations as separate clusters. **[Marks:10]**
 - Stage 2: Inner/Outer Separation: For each of the three clusters identified in Stage 1, apply another clustering algorithm to further divide the points into "inner" and "outer" parts of that circle. **[Marks:15]**
- (b) Visualize the results: Plot the original data points, color-coded according to the final clusters obtained after each stage. The visualization for this part should be saved as `roll_no_figure1.png` and `roll_no_figure2.png`. **[Marks:15]**
- (c) Discuss your choice of clustering algorithms for each stage and the reasoning behind your parameter selection. **[Marks:10]**
- (d) Treat the obtained cluster labels from the two-stage approach as the target variable (y) and the original (x, y) coordinates as the features (X). Split your data into training and testing sets (e.g., 70% train, 30% test). Choose a supervised classification algorithm and train your chosen classifier using the training data. Evaluate the performance of your trained classifier on the testing data. Save the visualization as `roll_no_figure3.png`. **[Marks:20]**
- (e) Implement a clustering approach in a single shot to directly identify the six clusters (inner and outer parts of the three circles). Apply a suitable clustering algorithm with the appropriate number of clusters. Visualize the results. Save the visualization of the results for this part as `roll_no_figure4.png`. **[Marks:10]**